Should we break a Wireless Network into Sub-networks?*

Amir Dana†, Radhika Gowaikar†, Babak Hassibi†, Michelle Effros†,
Muriel Médard‡and Ralf Koetter §

Abstract

In this paper we show that to achieve capacity of a wireless network, the global structure of the network should be considered. In other words, achieving capacity on the sub-networks of a wireless network does not guarantee achieving capacity globally. We illustrate this fact by some examples. Then we consider packet erasure wireless networks with limited sets of operations allowed at each node. We pose some interesting problems related to the optimal achievable rate of these networks and provide partial answers to some of them.

1 Introduction

In network problems that involve routing, relays, multihop communication etc. a standard assumption is that each transmission is error-free. In other words, each link is operated below (or at) its capacity to ensure error-free transfer of information. In addition to this, in a wireline network, a node can send out different signals on each outgoing link and there is no interference among the various signals coming in to a node on different incoming links. In wireless networks, however, each node is forced to broadcast the same message on each outgoing link and interference between the different signals coming in at a node is (typically) unavoidable.

A type of max-flow min-cut theorem for general multiterminal networks is given in [1]. Using this theorem one can find different upper bounds for achievable rates in arbitrary networks. But, as mentioned in [1], these bounds are not always tight even for simple channels.

In the wireline setup, for a family of network problems known as multicast problems, it has been shown that the max-flow min-cut upper bound can actually be achieved [2], [3], [4]. In this problem the network consists of a source, destinations and relay nodes. Each of the links between two nodes represents a memoryless channel with some known

*This work was supported in part by the National Science Foundation under grant no. CCR-0133818, by the office of Naval Research under grant no. N00014-02-1-0578, and by Caltech's Lee Center for Advanced Networking.

†Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125. email: {amirf,gowaikar,hassibi,effros}@caltech.edu

‡Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139. email: medard@mit.edu

§Coordinated Science Laboratory, Universtiy of Illinois, Urbana, IL 61801. email: koetter@uiuc.edu

capacity. All channels are assumed to be independent. The goal is to convey all the information from the source to each destination reliably.

The work of [2],[3] shows that by using channel coding on each link, so as to make the links error-free, and by employing network coding at the nodes, the max-flow min-cut upper bound is achievable. The authors further show that routing is not sufficient to reach the max-flow min-cut capacity. They propose a linear coding scheme over the network which permits nodes to send out different linear combinations of the incoming data on each outgoing link. This idea is put into an algebraic framework in [4] and applications of this framework appear in [4], [5].

Thus, for the multicast problem in wireline networks one can achieve the max-flow min-cut capacity using linear network coding separated from and on top of the channel coding performed for each link in the physical layer of the network. By using channel coding and decoding on each link one can pose the problem on an error-free network without losing performance; and by means of linear network coding one can achieve the capacity.

In view of these results for wireline networks, one interesting and important question that arises is whether the same approach gives satisfactory results for multicast over wireless networks. We know that for the multicast problem over wireline networks making each link error-free does not degrade the performance of the network. Therefore one can decompose the network into smaller modules. The question is whether this is also true for wireless networks.

As we will see in this paper, for wireless networks this approach is not optimal and can cause severe degradation in the performance of the entire network. In other words, for optimal operation of the network, its global structure should be considered and a coding scheme for the entire network, rather than for each individual link is required. This fact is noted in [6], [7] as well.

The paper is organized as follows. In the next section, we provide some examples to illustrate that viewing a wireless network as a composition of sub-networks and designing them separately is not necessarily the best thing to do. In section 3, we introduce our network model, which is a network of erasure channels with broadcast. We will also specify the set of operations allowed at each node. To gain insight into this class of networks, section 4 studies some further examples and derives their optimal rates and policy. Section 5 poses some important questions and attempts to provide partial answers. Section 6 is the conclusion.

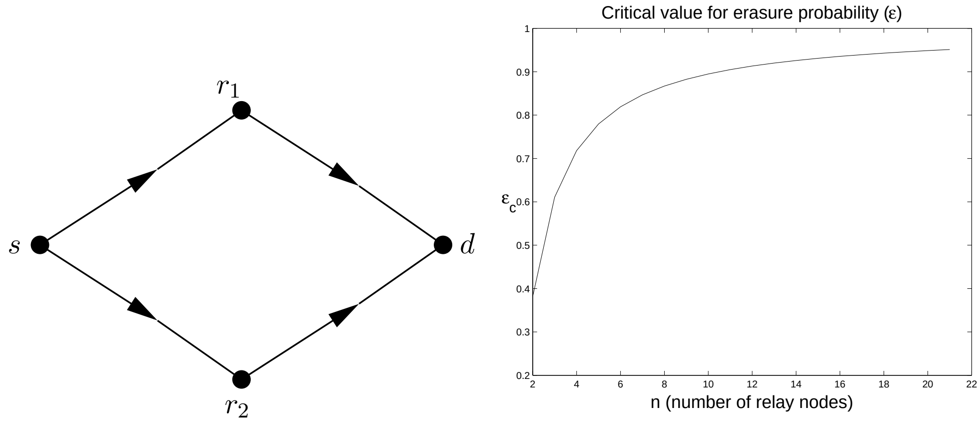
2 Some Examples

In this section we consider two different for which decoding at intermediate nodes is not always optimal. Some other examples can be found in [6].

2.1 Gaussian Relay Networks

The first network is a Gaussian parallel relay network consisting of 2 relay nodes and one source-destination pair (See Fig.1.(a)). The relay nodes r_1 and r_2 are solely to aid communication from source to destination. We assume that the noise power at each receiver is σ_n^2 and the transmit power at each node is p.

If one views the network as a broadcast channel cascaded with a multi-access channel (in other words, if we assume that the relays decode their messages correctly and code



(a) (b)

Figure 1: (a) Graph representation of a relay network with 2 relay nodes, (b) Critical value for erasure probability as a function of the number of nodes in the relay network.

them again and transmit) then the capacity of the network is bounded by the sum-rate capacity of the broadcast channel. Since the Gaussian broadcast channel is degraded, the sum-rate capacity is known and is equal to

$$R_d = \log(1+\rho)$$

where $\rho \triangleq \frac{p}{\sigma_n^2}$ is the Signal to Noise ratio (SNR) and the subscript d in R_d implies that R_d is the maximum rate achievable by decoding at relay nodes. Now consider another strategy in which the relay nodes do not decode but only normalize their received signal and transmit it to the destination. In this case the received signal at the destination is

$$y = \sqrt{\frac{p}{p+\sigma_n^2}}(2x + n_1 + n_2) + n_3$$

where x, y, n_1, n_2, n_3 are the transmitted signal from the source, the received signal at destination and the noises introduced at relay nodes and the destination respectively. For this case the received signal is a scaled version of x added with Gaussian noise. The maximum achievable rate, denoted by R_f, is

$$R_f = \log(1 + \frac{\frac{4p^2}{p+\sigma_n^2}}{\sigma_n^2 + \frac{2p\sigma_n^2}{p+\sigma_n^2}})$$

By simplifying this equation and writing it in terms of ρ we have

$$R_f = \log(1 + \frac{4\rho^2}{3\rho+1})$$

Comparing this to the achievable rate when the relay nodes "decode and re-encode", we can see $\rho = 1$ is a critical value in the sense that if $\rho > 1$ then we have superior performance in the "forward"ing scheme. If $\rho < 1$, then we have better rate when we "decode and re-encode" at relay nodes. If we increase the number of relay nodes from 2 to n (thus increasing the number of paths from source to destination), it can be easily checked that this critical value moves toward zero for Gaussian relay networks. Therefore in the limit of $n \to \infty$ it is always favorable to "forward". This fact is also true for Gaussian relay networks in the presence of fading. The work of [7] shows that for fading Gaussian relay networks with n nodes, the asymptotic capacity achievable with the "decode and re-encode" scheme scales like $O(\log\log n)$ whereas with the "forward" scheme it scales like $O(\log n)$. We should remark that decoding at one of the relay nodes and forwarding at the other is always sub-optimal.

2.2 Erasure Relay Networks

In this example, we consider a relay network with two relay nodes and the same graph representation as in Fig.1.(a). Assume that each of the links is now an erasure channel with probability of erasure $0 < \epsilon < 1$ independent of other channels. Furthermore we assume that the channel between the relay nodes and the receiver is a multiaccess channel with no interference. The channel between the source and the relay nodes is assumed to be a broadcast channel. As in the previous example we can assume that the relay node can "decode and re-encode" or just "forward". If the relay nodes choose to decode then the capacity of the network is bounded by the sum-rate capacity of the erasure broadcast channel. Since the erasure broadcast channel is degraded the maximum sum-rate is known and is equal to

$$R_d = 1 - \epsilon$$

where R_d denotes the maximum achievable rate when relay nodes are decoding. If the relay nodes just forward what they have received, the probability of erasure for the entire network is the probability that in both paths from source to destination (each passing through one relay node) at least one erasure occurs. Thus the maximum achievable rate, denoted by R_f, is

$$R_f = 1 - (1 - (1 - \epsilon)^2)^2$$

It can be verified that for values of $\epsilon < \frac{3-\sqrt{5}}{2}$ it is better to "forward" rather than to "decode and re-encode" at relay nodes. (We should again remark that decoding at one of the relay nodes and forwarding at the other is always sub-optimal.)

As in the Gaussian relay networks we can see that if the quality of the links in the network is better than some critical value, i.e. if $\epsilon < \frac{3-\sqrt{5}}{2}$, it is better to "forward" rather than do "decode and re-encode". Also if we increase the number of relay nodes from 2 to n, it can be easily checked that this critical value becomes closer to one. Therefore in the limit of $n \to \infty$ it is always favorable to "forward". In Fig. 1.(b) we have plotted the critical value of erasure probability, $\epsilon_c(n)$ for different values of n. We observe from the plot that $\epsilon_c(n)$ is an increasing function of n with limiting value of one as $n \to \infty$.

2.2.1 Another Possibility: "Encode"

In the example above, we considered only two operations allowed at every node: "decode and re-encode" and "forward". If we assume that packets have headers (i.e. they are numbered) and the overhead of numbering them is not significant, we can allow for another operation, which we call "encode". Using "encode" at relay nodes we can achieve a capacity of $1-\epsilon^2$ for the erasure relay network in Fig.1.(a). To see this, suppose that we have $n(1-\epsilon^2)$ packets of data at the source to convey to the destination. We can encode them to n packets using channel codes so that if the destination receives $n(1-\epsilon^2)$ of these without erasure it can retrieve all the data. Sending the encoded packets through the network, each of the relay nodes will receive $n(1-\epsilon)$ of the packets with high probability (assuming n is large enough). Now since each packet has a header there is no need for the relay nodes to send out an erasure packet for the erased packets. Moreover, since the capacity of the channel between each of the relay nodes and the destination is $1-\epsilon$, each of the relay nodes can encode the $n(1-\epsilon)$ packets they have received to n packets and transmit it across the channel so that the destination can retrieve the $n(1-\epsilon)$ packets with arbitrarily small probability of error. Since the packets have headers the destination knows which packets have been erased. Thus the communication between any relay node and the destination becomes error-free with this encoding. Looking at the network now, we have two parallel link from source to destination each with probability of erasure equal to ϵ. Thus the destination will receive $n(1-\epsilon^2)$ of the packets transmitted by the source with high probability. Using these, the destination can decode the message to arbitrary accuracy.

These observations suggest that in order to design an efficient wireless network code one cannot consider only its local structure the total structure of the network needs to be taken into account.

3 Network Model

In wireless networks the medium is shared between all the users. There is no notion of a "link" in a wireless network, but by specifying a protocol for communication between the nodes, the network can be represented by a directed graph. All the outgoing edges from a node carry the signal transmitted from that node. We assume that all the transmissions are simultaneously done.

In this work we consider packet erasure networks for analysis. The reason for this choice is that this model seems reasonable for analyzing packet switched networks over wireless channels. Also, this problem seems to be more tractable than that of a wireless network with Gaussian links.

For the reception at nodes, we assume that by using some division multiple access scheme (for instance, CDMA, TDMA or FDMA) each node can receive all the incoming signals without interference among them. Thus the network is wireless only in a broadcast sense.

The possible operations at each node are restricted to the following three:

- **"Forward"**: In this case a node simply transmits the packets it has received (from at least one of its incoming edges) and erasures in place of the other packets. In other words, it "forwards" all that it receives.

- **"Decode and re-encode"**: In this case a node can decode the received packet exactly and determine the message. Then it transmits the codeword corresponding to the message using the same codebook as that of the source. In short, it transmits exactly what the source transmitted. We should remark that performing "Decode and re-encode" at a node introduces a constraint on the rate of communication because we are assuming that the node can decode the message correctly. On the other hand, when a node decodes it can be viewed as a copy of the source for all its successor nodes.

- **"Encode"**: This operation is possible only if we assume that packets have headers (numbers) and can be distinguished from each other. Suppose that a node v receives a fraction of the total number of packets sent by the source. If this fraction is less than the capacity of all the edges going out from v then v can encode its received packets so that its immediate successors can retrieve those packets without error. This is similar to the example in section 2.2.1. In other words, outgoing links from v seem noise-free. We should remark that the immediate successors of v should first decode the packets encoded by v and then perform one the possible operations.

Of course there are some limitations concerning this operation. If we assume that packets have headers then part of the packet does not convey information and only designates the header. As the number of packets increases we need more bits for numbering the packets. Therefore we should assume that the size of packets is large enough so that the header size is negligible. But as we saw in section 2 using this scheme, one may improve the performance significantly.

3.1 Problem Statement

The network is represented by an acyclic directed graph (V, E) with one source-destination pair $(s = v_0, d = v_{n+1})$. $V = \{v_0, v_1, v_2, \ldots, v_n, v_{n+1}\}$ denotes the set of vertices and E denotes the set of edges (links) in the network. Each edge is an ordered pair from $V \times V$. Associated with edge (v_i, v_j) is the probability of erasure on that edge ϵ_{ij}. A packet sent from node v_i is received correctly with probability $1 - \epsilon_{ij}$. We further assume that the erasure events are independent for all the edges and the links are memoryless. Therefore every network is completely characterized by $(V, E, \{\epsilon_{ij}\})$.

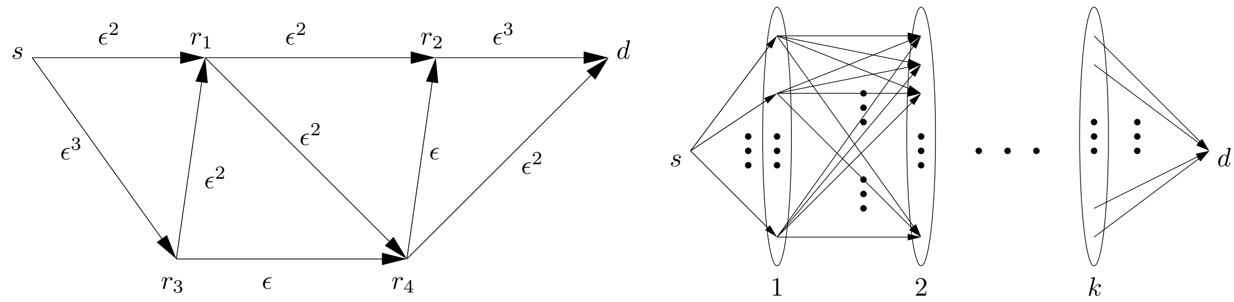
The problem is to find the optimal rate and the policy that results in the optimal rate. Each policy is identified by $D = (D_d, D_e), \quad D_d, D_e \subseteq V - \{s, d\}, D_d \cap D_e = \emptyset$. D_d is the subset of relay nodes that performs the "decode and re-encode" operation and D_e is the subset of relay nodes that performs "encode" operation. Nodes in $V - \{s, d\} - (D_e \cup D_d)$ "forward" their data. A given policy D is admissible if the fraction of the packets that each node v in D_e receives (in other words the rate of transmission of data to v) is less than the capacity of any of its outgoing edges.

One solution to find the optimal rate and the optimal policy for a given network is to search over all admissible policies and find the one with the best performance. The complexity of this search is at least exponential in the number of nodes in the network. An important problem is to propose an algorithm that finds the optimal rate in a more efficient manner.

For simplicity, in the rest of this paper we address a more restricted version of the above problem. We optimize the performance of the network over the policies in which no node performs "encoding", i.e. a node can only "forward" or "decode and re-encode". Since each of the relay nodes has two choices, the number of policies is $2^{|V|-2}$. It is clear that in this case all $2^{|V|-2}$ policies are admissible. Therefore the rate is simply specified through D_d. We refer to this as D in the rest of the paper.

3.2 Policy and Rate

For each $v \in V$ let $e_D(v)$ denote the probability that v does not receive the data packet under policy D. Clearly $e_D(v)$ is the probability that on every path from $D \cup \{s\}$ to v there is at least one link on which erasure occurs. We can formalize this quantity by assigning each link (v_i, v_j) in E an independent Bernoulli random variable b_{ij} where $\Pr(b_{ij} = 0) = \epsilon_{ij}$. This random variable is zero if erasure happens on link (v_i, v_j), otherwise it is equal to one.



(a) (b)

Figure 2: Two examples of erasure networks (a) A network with four relay nodes (b) A mutistage relay network with k stages

Consider a path from a node $v' \in \{s\} \cup D$ to an arbitrary node v. Associate with this path the Boolean expression formed by performing logical "AND" on all the b_{ij}'s corresponding to the edges on that path. Therefore the connection between v' and v through that specific path is broken iff the Boolean expression corresponding to that path is equal to zero. Only when the connection from every $v' \in \{s\} \cup D$ to v is broken does v receive an erasure. Therefore

$$\begin{aligned} e_D(v) &= \Pr\{b_{v'i_1} \wedge b_{i_1 i_2} \wedge \cdots \wedge b_{i_j v} = 0 | \; \forall \text{ paths } (v', i_1, \ldots, i_j, v) \; ; v' \in D \cup \{s\}\} \\ &= \Pr\{\cup_{(v', i_1, \ldots, i_j, v); v' \in D \cup \{s\}} (b_{v'i_1} \wedge b_{i_1 i_2} \wedge \cdots \wedge b_{i_j v} = 0)\} \\ &= \Pr\{\vee_{(v', i_1, \ldots, i_j, v); v' \in D \cup \{s\}} b_{v'i_1} \wedge b_{i_1 i_2} \wedge \cdots \wedge b_{i_j v} = 0\} \end{aligned}$$

where $\wedge$ and $\vee$ denote the logical "AND" and "OR" respectively. Thus we know how to find $e_D(v)$ for every node.

The maximum achievable rate for the network under policy D is denoted by R_D. As mentioned earlier, performing "decode and forward" at each node of set D introduces some constraints on the maximum achievable rate of the network. Clearly, in order to be able to decode the original message perfectly at some node v', the rate of information transmission should at most be equal to $1 - e_D(v')$. Therefore the maximum achievable rate is equal to

$$R_D = 1 - \max\{e_D(v) | v \in \{d\} \cup D\}$$

4 Some More Examples

4.1 A Network with Four Relays

Consider the network shown in Fig. 2.(a). In this network there are four relay nodes. Erasure probabilities are shown on each link and are powers of ϵ. We are interested in the optimal rate and the corresponding policy for different values of ϵ. One way to find the maximum rate for different values of ϵ is by searching over all the 2^4 possible policies. (Another more efficient algorithm will be proposed in section 5.)

Fig. 3.(a) shows the rate as a function of ϵ. This plot is generated by exhaustive search over all the possibilities. We find that there are five different policies for different values of ϵ. Policies $D = \emptyset$, $D = \{r_1\}$, and $D = \{r_1, r_2, r_3, r_4\}$ give the optimal rate for intervals $[0, 0.2325]$, $[0.2325, 0.5785]$, $[0.5785, 1]$ respectively.

4.2 Multi-stage Relay Network

In the second example consider a multi-stage relay network as depicted in Fig. 2.(b). In this network there are k intermediate stages of relay nodes. $l = (l_1, \ldots, l_k)$ specifies the number of relay nodes in different stages. As the figure indicates, any two adjacent stages are fully connected. In our example we consider a network with 11 stages and $l = (4, 7, 6, 8, 8, 7, 6, 10, 4, 7, 5)$. We further assume that the probability of erasure on each link is ϵ.

Note that because of the symmetry among the nodes on the same stage, if one node v in stage m decodes in the optimal policy, without loss of generality we can assume that all the nodes in stage m decode as well. By searching over all the 2^{11} policies, we have found the optimal rate and the optimal policy for different values of ϵ. (Another more efficient algorithm is proposed in [8] that is linear in the number of layers rather than exponential in the number of layers.) The optimal rate is plotted in Fig. 3.(b). As we see there are two regions in the plot. In the linear region, the nodes in the first relay stage should "decode and re-encode". The operation of all the other stages does not change the rate. Thus there is more than one policy that gives the optimal rate. In the nonlinear region of the curve, nodes in stage 9 should always "decode and re-encode" and those in stages 1 to 8 should "forward".

5 Questions and Partial Answers

There are many important questions related to the wireless networks as modeled in this paper. In this section we pose some of these questions and problems. Providing answers to these questions is part of an independent work [8]. In this paper, we only state some of these results as theorems without proof. For derivation and further details we refer the reader to [8]. Consider an erasure network $(V, E, \{\epsilon_{ij}\})$.

- **Problem 1** Find simple upper and lower bounds for the capacity of the network based on its structure. Also determine how tight these bounds are.

  A type of max-flow min-cut upper bound can be found using the following definition.

  Definition Given a network $(V, E, \{\epsilon_{ij}\})$, a source/destination cut denoted by $[S, T]$ is a partition of the nodes of the network into a source set S containing source node s and a destination set T containing the destination node d. The value of a source/destination cut, $[S, T]$, is defined to be the product of the probabilities of erasures on the links corresponding to the edges from S to T. This value is denoted by $\text{Val}_{cut}([S, T])$.

  Theorem 1. *Consider a wireless erasure network $(V, E, \{\epsilon_{ij}\})$ with operations "forward" and "decode and re-encode", allowed at every node. The capacity of the network, C, is always upper bounded by one minus the maximum value of all the cuts in the network. i.e.,*

  $$C \leq 1 - \max\{\, Val_{cut}([S, T]) | [S, T] \text{ is a source/destination cut}\}$$

  Corollary 1. *For a given network $(V, E, \{\epsilon_{ij}\})$ with $\epsilon_{ij} = \epsilon_0$ for all the links, the capacity is bounded by*

  $$C \leq 1 - \epsilon_0^{\kappa'(s,d)}$$

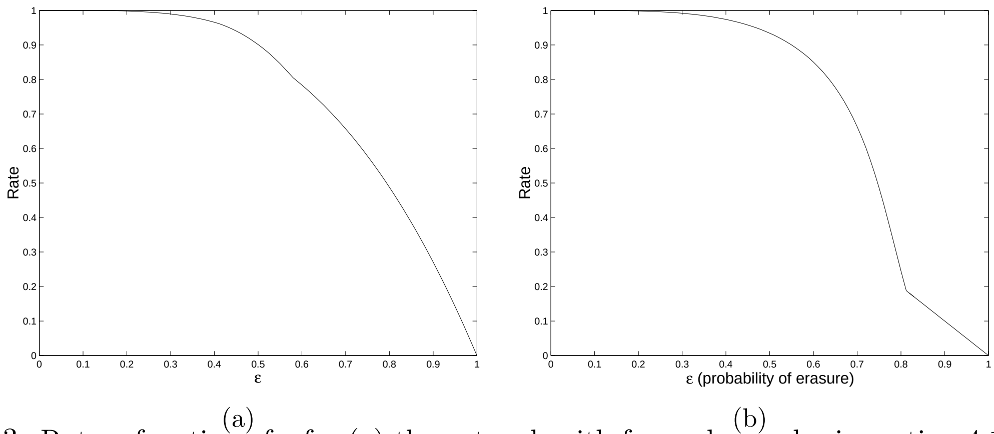


Figure 3: Rate a function of ϵ for (a) the network with four relay nodes in section 4.1 and for (b) multistage relay network with 11 stages in section 4.2

where, as defined in [9], $\kappa'(s,d)$ is the minimum size of a source/destination cut (or the size of a min-cut).

- **Problem 2** Propose an efficient algorithm that finds the optimal rate and the corresponding policy for a network.

  As we saw in the previous section, for large networks exhaustive search is too time consuming. The following theorem proposes an efficient algorithm for finding the optimal rate when the nodes are restricted to the "forward" and "decode and re-encode" operations.

  Theorem 2. *For any erasure wireless network $(V, E, \{\epsilon_{ij}\})$ as modeled in section 3, the following algorithm gives the optimal rate and provides a policy to achieve this rate.*

  1. *Set $D = \emptyset$.*
  2. *Find $e_D(v)$ for all $v \in V$.*
  3. $D' \triangleq \operatorname{argmin}_{v \in Q} e_D(v) \quad Q = \{v | v \in V - (D \cup \{s,d\}), e_D(v) \leq 1 - R_D\}$
  4. *If $D' = \emptyset$ go to step 6. Otherwise $D = D \cup D'$.*
  5. *Go to step 2.*
  6. *$D_{opt} = D$ and $R_{opt} = R_D$*

In the remaining problems we assume that the probability of erasure on all the links is equal, i.e. $\epsilon_{ij} = \epsilon_0$.

- **Problem 3** The value of ϵ_0 shows the quality of the links in the network. If ϵ_0 is close to 1, it indicates that, with high probability, erasures will occur. At the other extreme, ϵ being near zero indicates that the links are reliable. It is interesting to see how the optimal rate behaves in these two asymptotic cases ($\epsilon_0 \to 0$ and $\epsilon_0 \to 1$) and determine what policy gives that optimal rate.

- **Problem 4** As we see from the examples in the previous sections the policy that gives the optimal rate is different for different values of ϵ_0. Therefore there are some critical values of ϵ_0 at which the optimal policy changes. Characterization of these critical values and the optimal policy for different intervals of ϵ_0 is an interesting problem.

There are also some more fundamental questions that are very much open.

- **Problem 5** We assume that by means of some division multi-access technique each node receives each of the incoming signals perfectly without interference among them. Coming up with some model for multi-access channels in erasure networks that incorporates the interference yet is still amenable to analyses is of great interest.

- **Problem 6** The set of operations allowed at each node is limited in our model. Therefore the optimal rate achieved may not necessarily be the capacity of the network. Proving that the capacity is indeed the optimal rate achieved in our problem or coming up with optimal operations at each node that achieve the capacity of the network is an open problem.

6 Conclusion

In this paper we showed that the traditional ideas regarding the design of wired networks cannot always be used in the wireless setting. By using some examples we showed that designing the sub-networks of a wireless network optimally does not necessarily imply good performance for the total network. For optimal operation of the network one should look at its complete structure. Motivated by this, we considered a special class of wireless networks, namely, erasure wireless networks for further analysis. We formalized the problem and posed a number of interesting questions. Partial answers to some of them were mentioned briefly.

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, New York, 1991.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Info. Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[3] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Info. Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[4] R. Koetter and M. Médard, "Beyound routing: and algebraic approach to network coding," *Proceeding of INFOCOM 2002*, vol. 1, pp. 122–130, 2002.

[5] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros, "The benefits of coding over routing in randomized setting," *Proceeding of the ISIT 2003*, vol. 1, pp. 122–130, 2003.

[6] M. Effros, M. Médard, T. Ho, S. Ray, D. Karger, and R. Koetter, "Linear network codes: a unified framework for source channel, and network coding," *invited paper to the DIMACS workshop on network info. theo.*, 2003.

[7] A. F. Dana, M. Sharif, B. Hassibi, and M. Effros, "Is broadcast plus multiaccess optimal for Gaussian wireless networks with fading?," *to appear in the 37 Asilomar Conf. on Sig., Sys. and Comp.*

[8] A. F. Dana, R. Gowaikar, B. Hassibi, and M. Effros, "On the capacity of erasure wireless networks," *in preparation for submission.*

[9] D. B. West, *Introduction to graph theory*, Prentice Hall, 1996.